Tutorial

Image Classification

with

TNTmips®

Before Getting Started

This tutorial booklet introduces you to the TNTmips® procedures for automatic classification of multispectral and multi-temporal imagery. The process includes "unsupervised" methods, which automatically group image cells with similar spectral properties, and "supervised" methods, which require you to identify sample areas. Automatic statistical analysis of the classes helps you interpret the results and guides you through optional interactive merging of similar classes.

Prerequisite Skills This booklet assumes that you have completed the exercises in the tutorial booklets entitled *Displaying Geospatial Data* and *TNT Product Concepts*. Those exercises introduce essential skills and basic techniques that are not covered again here. Please consult those booklets for any review you need.

Sample Data The exercises presented in this booklet use sample data that is distributed with the TNT products. If you do not have access to a TNT products DVD, you can download the data from MicroImages' web site. In particular, this booklet uses objects in the RGBCROP Project File in the CROPDATA data collection, the CB_TM Project File in the CB_DATA data collection, and the BERMNDVI and BEREATRN Project Files in the BEREA data collection.

More Documentation This booklet is intended only as an introduction to the Automatic Classification process. Details of the processes discussed can be found in a variety of tutorial booklets, Technical Guides, and Quick Guides, which are all available from MicroImages' web site.

TNTmips® Pro and TNTmips Free TNTmips (the Map and Image Processing System) comes in three versions: the professional version of TNTmips (TNTmips Pro), the low-cost TNTmips Basic version, and the TNTmips Free version. All versions run exactly the same code from the TNT products DVD and have nearly the same features. If you did not purchase the professional version (which requires a software license key) or TNTmips Basic, then TNTmips operates in TNTmips Free mode. All the exercises can be completed in TNTmips Free using the sample geodata provided.

Randall B. Smith, Ph.D., 21 April 2011

You can print or read this booklet in color from MicroImages' web site. The web site is also your source for the newest tutorial booklets on other topics. You can download an installation guide, sample data, and the latest version of TNTmips Free.

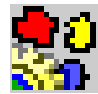
http://www.microimages.com

Welcome to Image Classification

Many remote sensing systems record brightness values at different wavelengths that commonly include not only portions of the visible light spectrum, but also photoinfrared and, in some cases, middle infrared bands. The brightness values for each of these bands are typically stored in a separate grayscale image (raster). Each ground-resolution cell in an image therefore has a set of brightness values which in effect represent the “color” of that patch of the ground surface, if we extend our concept of color to include bands beyond the visible light range.

The Automatic Classification process in TNTmips uses the “colors”, or **spectral patterns**, of raster cells in a multispectral image to automatically categorize all cells into a specified number of spectral classes. The relationship between spectral classes and different surface materials or land cover types may be known beforehand, or determined after classification by analysis of the spectral properties of each class. The Automatic Classification process offers a variety of classification methods as well as tools to aid in the analysis of the classification results.

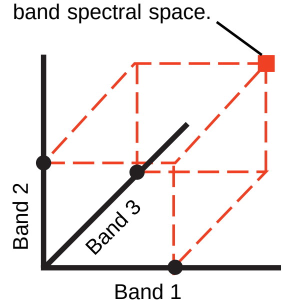
The spectral pattern of a cell in a multispectral image can be quantified by plotting the raster value from each band on a separate coordinate axis to locate a point in a hypothetical “spectral space”. This spectral space has one dimension for each band in the image. Most classification methods use some measure of the distance between points in this spectral space to assess the similarity of spectral patterns. Cells that are close together in spectral space have similar spectral properties and have a high likelihood of imaging the same surface features.

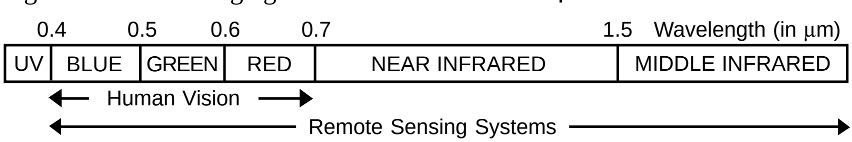


☑ select Image / Interpret / Auto-Classify from the TNTmips menu

Pages 4-12 lead you through several unsupervised classification methods and their user-defined parameters. Exercises introducing the tools for analyzing the classification result and merging classes are found on pages 13-19. Pages 20-26 introduce methods of supervised classification of images and use of the Error Matrix tool. The final series of exercises on pages 27-34 show you how to create training set and mask rasters with the Training Set Editor.

Location of a single spectral pattern in a three-band spectral space.





Unsupervised Classification

In **Unsupervised Classification**, TNTmips uses a set of rules to automatically find the desired number of naturally-occurring spectral classes from the set of input rasters. The rules vary depending on the classification method you choose from the Method option menu.

STEPS
- ☑ click the [Rasters...] button on the Automatic Classification window
- ☑ select raster objects RED, GREEN, and BLUE from the RGBCROP Project File in the CROPDATA data collection
- ☑ press the Method option menu button and select Simple One-Pass Clustering
- ☑ select None from the Input Redistribution option menu
- ☑ change the Number of Classes to 10
- ☑ choose Run from the File menu
- ☑ use the standard object selection procedure to create a new Project File named CROPCLAS for the output rasters
- ☑ name the output Class and Distance rasters using the default names provided
- ☑ click [OK] on the Select Object window to confirm the selection and start the classification process

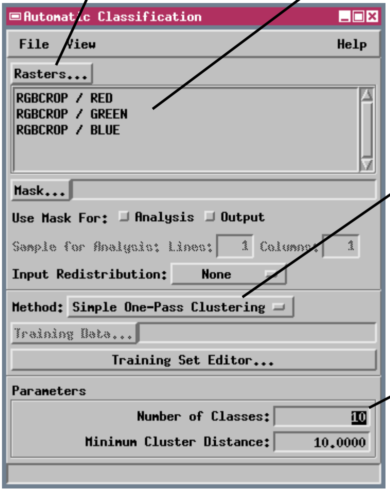
Keep the Automatic Classification window open with the current settings for the next exercise.

Let's begin by performing an unsupervised classification of the red, green, and blue raster components of a scanned, natural-color aerial photograph. (You will display this image for comparison with the classification results in a later exercise.)

Choose the Simple One-Pass Clustering Method for this exercise. This method establishes initial class centers and assigns cells to classes in one processing pass by determining the spectral distance between each cell and established class centers. Each raster cell is assigned to the nearest class; a cell too far away from existing class centers becomes the center of a new class (up to the specified number of classes).

Click the [Rasters...] button to select a set of input rasters for classification.

Selected input rasters are shown in the scrolled pane.



Click on the Method option button to reveal a menu of classification methods.

To change a parameter value, highlight the field with the mouse cursor and type in the desired value.

The classification method, number of classes, and other settings that you last use in a classification session are saved as default settings and are restored the next time you open the process. Be sure to check all settings before running a new classification.

Class and Distance Rasters

The main classification results are contained in the Class raster, which is automatically displayed in a View window. Some classification methods also give you the option of creating a Distance raster, which you can also select for viewing. Your output rasters should look like the illustrations below.

The **Class raster** is a categorical raster: each numerical value in the raster is an arbitrary class identifier that has been assigned to the cell by the classification process. The Unsupervised Classification process assigns class numbers in the order in which the classes are created. Because the raster values have no other numerical significance, for display a unique color is assigned to each class from a standard color palette.

The **Distance raster** is a grayscale raster that shows how well each cell fits its assigned class. Each raster cell value in the Distance raster records the distance between that cell and its class center in spectral space. Cells that are closer to their class center (better fit) appear darker in the displayed raster than those with greater distance values (poorer fit).

STEPS

☑ examine the Class raster in the View window

☑ from the Layer menu on the View window, select Controls

☑ in the Layer Manager window icon bar, click the Add Raster icon button and choose Single from the dropdown menu 

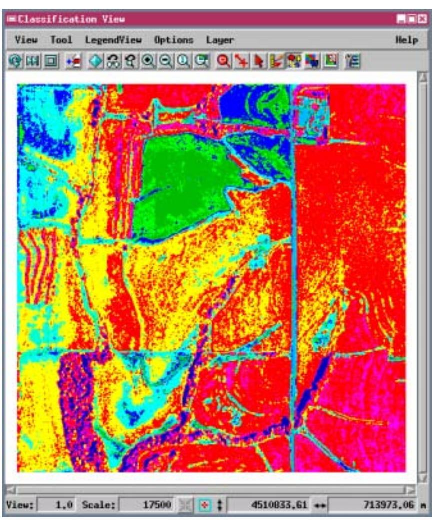
☑ select DST_SIMPLE from your CROPCLAS Project File

☑ click on the Layer Controls icon in the DST_SIMPLE layer entry 

☑ in the Raster Layer Controls window, select Auto Normalize from the Contrast option menu and click [OK]

☑ examine the Distance Raster

☑ right-click on the DST_SIMPLE entry in the Layer Manager and choose Remove Layer from the dropdown menu



Class raster CLS_SIMPLE (categorical raster)

Cells with poorer fit to their class have lighter tones in the Distance Raster.

Grayscale Distance raster DST_SIMPLE with auto-normalized contrast enhancement

Display the Input Image

STEPS

☑ choose Main / Display from the TNTmips menu

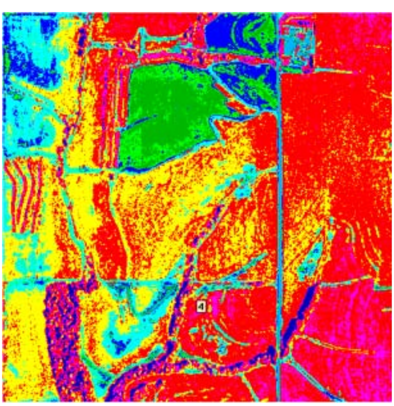
☑ in the Display Manager, choose Display / Open

☑ select GROUP1 from the RGBCROP Project File

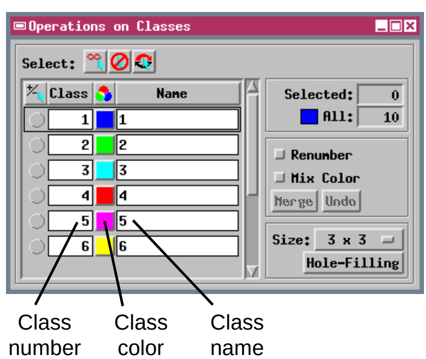
The input color image covers about 320 acres of farmland in eastern Nebraska. It was acquired during the summer growing season, and includes fields of corn, wheat, and soybeans. Our main purpose in classifying the image is to map the distribution of the different crops, but the scene also includes areas of trees and grass, a road, a pond, and farm buildings with bright roofs. The Automatic Classification process is influenced by the sets of brightness values of all cells in the scene, not just the features that you intend to classify.

RGB display of the three rasters used as input for the first set of classification exercises.

Arrange the Classification View and Group View 1 windows on the screen so that you can compare the Class raster to the input image. Note that most fields include cells of two or even three classes. The scrolled table in the Operations on Classes window shows the color assigned to each numbered class, and the class name (for unsupervised classification, the default name is the class number). In later exercises we will use this window to rename classes, merge similar classes, and simplify a class raster.



Class raster shown at the same scale.



☑ choose Display / Close from the Layer Manager window before proceeding to the next page

Compare with Ground Truth Data

In order to interpret the results of an unsupervised classification, it is useful to compare the Class raster to any available information about the types of materials and ground cover in the scene. Partial "ground truth" information for the RGBCROP airphoto is contained in two vector objects.

The CROPMAP vector object outlines field boundaries and has labels identifying the known cover types. Fields for which the crop is not known are labeled with a question mark. The STREAMS vector object outlines the drainage pattern of the area. The streams are flanked by narrow strips labeled "Trees", which actually include both trees and grassy areas.

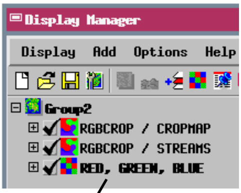
Note the range of classes found in the corn fields. The corn fields in the northwest corner of the image include classes 1, 3 and 6; the field in the center mostly classes 4 and 6; and those in the southeast corner mostly classes 4 and 5. This variation in spectral characteristics may result from differences in plant size, the density of the leaf canopy, soil type and conditions, slope direction, or other factors.

This variability is inherent in most of the land cover types that you may try to recognize in airphotos or satellite imagery. When you perform an unsupervised classification, you should therefore set the number of output classes to be several times greater than the number of land cover types that you hope to recognize. You can then use the analysis tools that are discussed in later exercises to recognize and group the similar spectral classes.

STEPS

☑ in the Display Manager, choose Display / Open

☑ select GROUP2 from the RGBCROP Project File



Display Manager window shown with "file name / object name" as layer names (to change this setting, choose Options / View Options from the Display Manager and use the Default Layer Name menu on the Layer tabbed panel).

The color photo overlaid with vector objects STREAMS (in blue) and CROPMAP (yellow polygons).

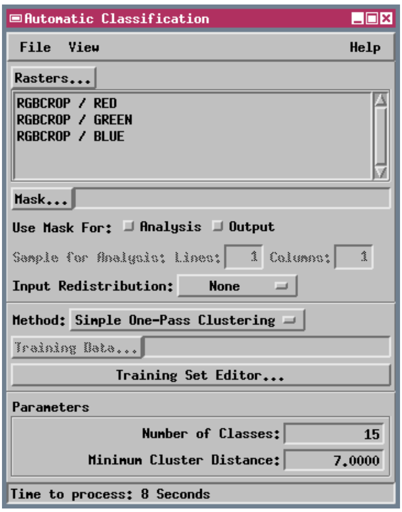
☑ choose Display / Close from the Layer Manager window before proceeding to the next page

Adjust Classification Parameters

STEPS
- ☑ change the Number of Classes parameter setting to 15
- ☑ change the Minimum Cluster Distance setting to 7.0
- ☑ run the classification process

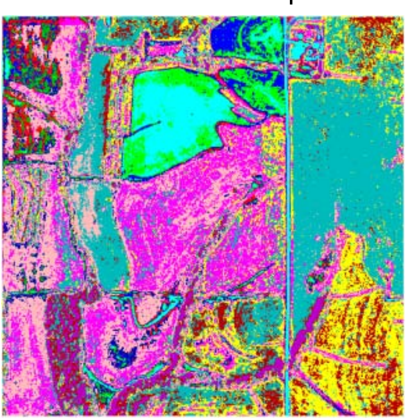
The Class raster that you produced in the exercise on page 6 does not differentiate between soybean and corn fields particularly well, since the same classes are found in both types of field. We can improve the classification by adjusting the parameters.

In the Simple One Pass Clustering classification method, the Minimum Cluster Distance parameter sets the threshold distance in spectral space used to designate an input cell as a new class center instead of assigning it to the closest class. By adjusting this parameter downward, you increase the likelihood that different land cover types that are close together in spectral space will be assigned to distinct classes.



The Number of Classes parameter sets an upper limit on the number of output classes. Increasing the output class limit also makes it more likely that similar cover types will be assigned to distinct classes rather than being lumped together in a single class.

The new Class raster shows a much better separation of corn, soybean, and wheat fields than the previous result. On the basis of this classification, the unknown fields in the southwest part of the image are probably corn fields. However, there is still substantial class overlap between the corn and soybean fields in the eastern part of the image. The easternmost unknown fields appear to be either corn or soybean fields, but it is impossible to assign them to either category with any confidence.



Class raster produced using the adjusted parameters, with 15 classes.

Apply a Mask During Classification

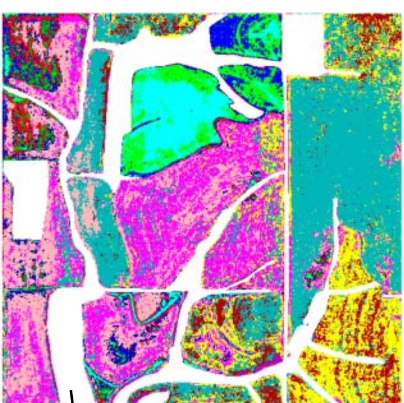
You can further fine-tune the classification of the RGBCROP photo by applying a mask during the classification process. In order to focus the classification process on differentiating crops, the FLD_MASK object masks out all of the tree-covered and other noncultivated portions of the RGBCROP image, as well as field edges shaded by adjacent trees. (You can create a mask raster for an input image using the Training Set Editor, which is introduced in later exercises.)

The Analysis toggle button applies the mask during the process of building classes, so that the set of classes is determined using only the unmasked areas. The Output toggle button restricts the final classification to the unmasked areas. Masked areas are assigned a value of 0 in the Class raster, and appear black. (Unlike other unsupervised classification methods, the One Pass Clustering method combines both operations in a single pass through the image.)

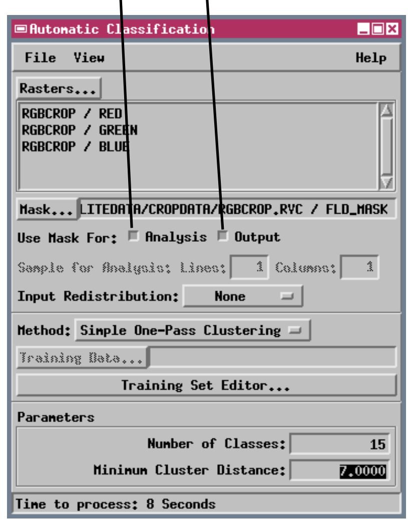
STEPS
- ☑ press the [Mask...] button on the Automatic Classification window
- ☑ select the FLD_MASK raster object from the RGBCROP Project File
- ☑ for the Use Mask For settings, turn on both the Analysis and Output toggle buttons
- ☑ run the classification process

A **mask** is a binary raster that restricts an operation to particular parts of the input image. Areas to be processed have a value of 1 in the mask, while areas to be excluded from processing have a value of 0.

Turn on the Analysis and Output toggle buttons.



Masked areas are assigned a value of 0 in the Class raster and are displayed in black.



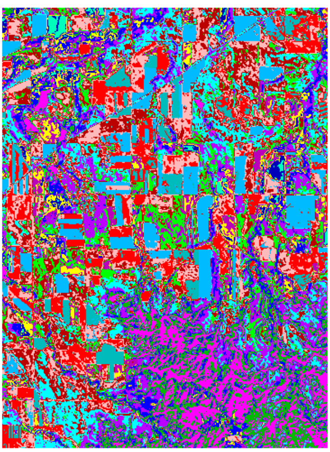
Fuzzy C Means Classification

STEPS

- ☑ from the CB_TM Project File in the CB_DATA data collection, select raster objects BLUE, GREEN, RED, PHOTO_IR, TM_5, and TM_7 for classification
- ☑ select Fuzzy C Means from the Method option menu
- ☑ in the Sample for Analysis text fields, change the Lines and Columns values to 2
- ☑ set the Number of Classes to 15, and the Maximum Iterations to 10
- ☑ run the classification process, and place the output Class raster in a new CB_CLASS Project File

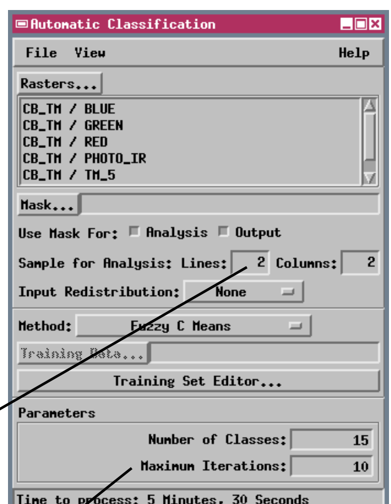
Although a color aerial photograph provides a familiar view of the landscape, better classification results can be achieved using multispectral imagery that incorporates infrared as well as visible light bands. The next series of exercises use a set of six Landsat Thematic Mapper bands as input for the classification process.

All of the unsupervised classification methods (except Simple One Pass Clustering) use a number of process cycles (iterations) to build classes from a sample of the input image cells before applying the classes to the entire image. The Fuzzy C Means classification method uses rules of fuzzy logic, which recognize that class boundaries may be imprecise or gradational. The Fuzzy C Means method creates an initial set of prototype classes, then determines a membership grade for each class for every cell. The grades are used to adjust the class assignments and calculate new class centers, and the process repeats until the iteration limit is reached.



Fuzzy C Means Class Raster.

Sample cells are selected at regular intervals throughout the image. Set the sampling intervals in the Line and Column directions using the Sample for Analysis text fields. With the default settings of 1, all input cells are used to build classes. Increasing these values speeds processing for large images. For example, changing both intervals to 2 results in a sample set made up of one quarter of the image cells.



The Maximum Iterations parameter sets an upper limit on the number of iterations performed in the class-building phase of the process.

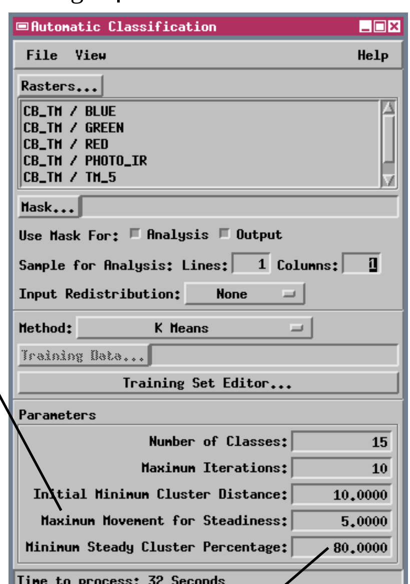
K Means Classification

The Automatic Classification process includes a number of unsupervised classification methods. Continue exploring these methods by applying the K Means classification method to the CB_TM raster set you used in the previous exercise.

The K Means algorithm executes more quickly than the Fuzzy C Means method, so we can conveniently use the entire image to build classes. The K Means method analyzes the input raster set to determine the location of initial class centers. In each process iteration, cells are assigned to the nearest class and new class centers are calculated. The new class center is the point that minimizes the sum of the squared distances between points in the class and the class center. With each iteration class centers shift and the class assignments for some cells change. The process repeats until the shift in class centers falls below a specific value or the maximum number of iterations is reached.

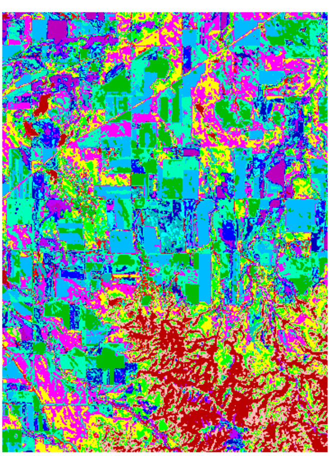
STEPS

☑ select K Means from the Method option menu
☑ change the Sample For Analysis text field values to 1
☑ if the Automatic Classification window does not automatically resize to show the complete parameter list shown below, resize the window manually
☑ accept the default settings for the other parameters
☑ run the classification process, and place the output class raster in the CB_CLASS Project File



A class center is considered to be steady when its movement with successive iterations falls below the Maximum Movement for Steadiness parameter value.

The Minimum Steady Cluster Percentage sets the percentage of class centers that must become steady in order to accept the current set of classes.



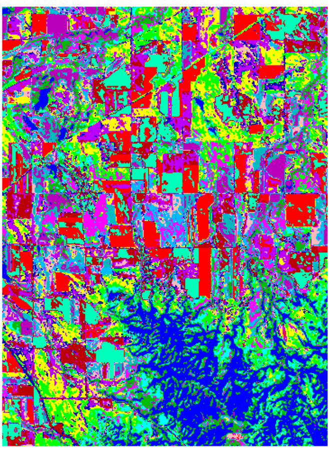
K Means Class raster for the CB_TM scene

ISODATA Classification

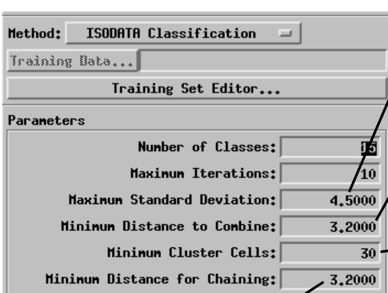
STEPS
- ☑ select ISODATA from the Method option menu
- ☑ resize the Automatic Classification window if needed to show all Parameters
- ☑ accept the current default settings for the parameters
- ☑ run the classification process, and place the output class raster in the CB_CLASS Project File

As a final example of unsupervised classification, apply the ISODATA method to the currently selected CB_TM raster set. The ISODATA method is similar to the K Means method but incorporates procedures for splitting, combining, and discarding trial classes in order to obtain an optimal set of output classes.

The ISODATA method determines an initial set of trial class centers and assigns cells to the closest class center. In each subsequent iteration the process first evaluates the current set of classes. A large class may be split on the basis of its number of cells, its maximum standard deviation, or the average distance of class samples from the class center. A class that falls below a minimum cell count threshold is discarded, and its cells are assigned to other classes. Pairs of classes are combined if the distance between their class centers falls below a threshold value. After classes have been adjusted, new class centers are calculated and the process repeats. Process iterations continue until there is little change in class center positions or until the iteration limit is reached.



ISODATA Class raster for the CB_TM scene



The Maximum Standard Deviation parameter provides one criterion for splitting large classes. If the class standard deviation for any input band exceeds this value, the class is split into two classes.

The Minimum Distance to Combine parameter sets the threshold distance used to determine if two nearby classes should be combined.

The Minimum Cluster Cells parameter sets the lower limit for the number of cells in a class. Any class with fewer cells is dissolved, and its cells are reassigned to other classes.

The Minimum Distance for Chaining parameter applies to the initial creation of class centers. It sets the lower limit on the distance between two class means.

Keep the Automatic Classification window open with the current settings for the next exercise.

View Output Statistics

The Automatic Classification process calculates several types of summary statistics for the final classes in each classification run. The statistics can be viewed in tabular form and in various graphical displays. This information enables you to investigate the spectral properties of each class and lets you compare classes for possible merging.

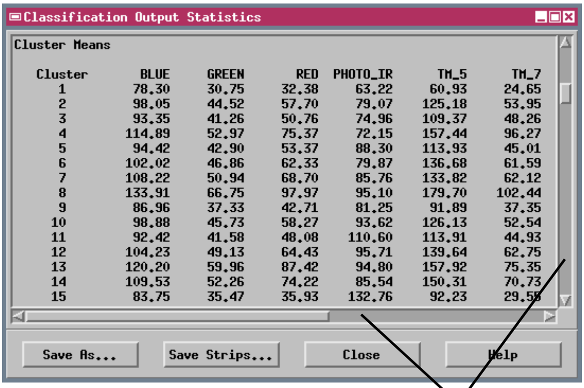
The Classification Output Statistics window tabulates Cluster Counts (the number and percentage of cells assigned to each class), Cluster Means, and Cluster Standard Deviations for each input band. The Cluster Distances between Means matrix lists the spectral distance between class centers for each pair of output classes. A Covariance Matrix for each class provides a relative measure of the degree of spectral correlation between each pair of input bands for that class.

STEPS

☑ select Output Statistics from the View menu on the Automatic Classification window

☑ click [Save As...] on the Classification Output Statistics window

☑ use the standard File / Object Selection procedure to name a text file in which to save the output statistics



Cluster	BLUE	GREEN	RED	PHOTO_IR	TM_5	TM_7
1	78.30	30.75	32.38	63.22	60.93	24.65
2	98.05	44.52	57.70	79.07	125.18	53.95
3	93.35	41.26	50.76	74.96	109.37	48.26
4	114.89	52.97	75.37	72.15	157.44	96.27
5	94.42	42.90	53.37	88.30	113.93	45.01
6	102.02	46.86	62.33	79.87	136.68	61.59
7	108.22	50.94	68.70	85.76	133.82	62.12
8	133.91	66.75	97.97	95.10	179.70	102.44
9	86.96	37.33	42.71	81.25	91.89	37.35
10	98.88	45.73	58.27	93.62	126.13	52.54
11	92.42	41.58	48.08	110.60	113.91	44.93
12	104.23	49.13	64.43	95.71	139.64	62.75
13	120.20	59.96	87.42	94.80	157.92	75.35
14	109.53	52.26	74.22	85.54	150.31	70.73
15	83.75	35.47	35.93	132.76	92.23	29.58

Use the scroll bars to scroll through the different statistics lists.

After you run the classification process, you may want to defer analysis of the classes to a later time. You can use the Open Class option on the File menu to select an existing class raster for analysis. The input bands are automatically reloaded, and all class statistics are calculated dynamically as needed. In addition, each statistics object can be saved as a file (a text file in the case of the Output Statistics.)

Cluster Distances between Means

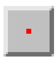
Cluster	1	2	3	4	5	6
2	80.363					
3	60.963	19.494				
4	134.922	59.557	76.356			
5	68.563	18.060	14.832	76.065		
6	95.384	15.293	34.463	45.459	31.890	
7	99.229	21.254	39.281	44.755	34.417	11.784
8	172.813	94.777	113.186	46.072	108.569	80.217
9	40.861	42.287	24.197	99.718	28.192	57.431
10	85.274	14.739	27.468	62.887	16.904	20.248
11	77.634	36.548	36.199	85.960	23.047	45.429
12	103.632	25.849	43.861	47.368	35.885	16.600
13	136.425	58.216	76.801	34.254	70.796	44.664
14	117.466	37.675	56.795	30.221	52.371	22.988
15	76.842	72.967	65.805	123.326	56.213	83.429

Choose Open Class to select an existing class raster for analysis.

Click [Close] to close the Classification Output Statistics window. Keep the Automatic Classification window open with the current settings.

Ellipse ScatterPlot

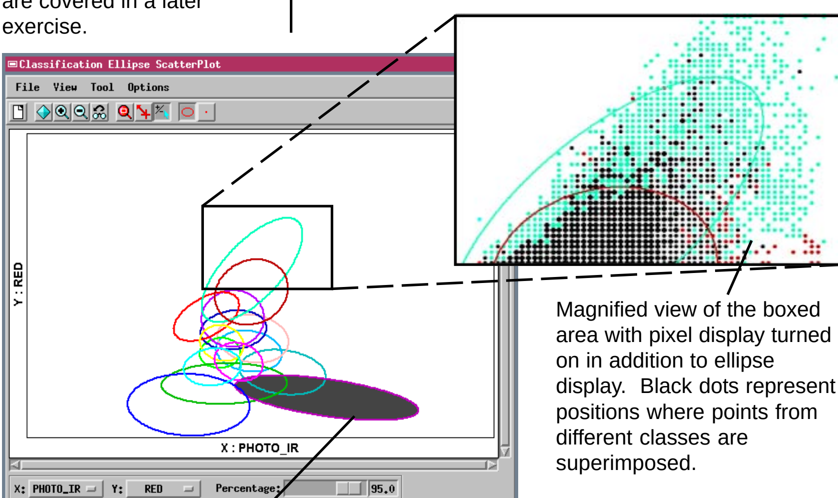
STEPS

- ☑ select Ellipse ScatterPlot from the View menu in the Automatic Classification window
- ☑ choose PHOTO_IR from the X option menu
- ☑ choose RED from the Y option menu
- ☑ click in the large, elongate ellipse that occupies the lower right part of the Ellipse ScatterPlot window
- ☑ press the Pixel icon button to turn on the point scatterplot viewing mode 

Classes selected in the Ellipse ScatterPlot window (or with the other graphical analysis tools discussed subsequently) are also simultaneously selected in the Operations on Classes window. Class operations are covered in a later exercise.

The Ellipse ScatterPlot window shows the distribution of classes projected onto a 2D plane through spectral space. You determine the orientation of the plane by choosing a pair of bands to assign to the X and Y axes of the diagram. In the default viewing mode each class is represented by an ellipse (drawn in the class color) that encompasses most of the scatter of points in the class. You can adjust the percentage of points included in the ellipse using the Percentage slider at the bottom of the window; the default value is 95 percent. You also have the option to view the scatter of cell values in addition to or in place of the ellipse display.

The positions of class clusters in spectral space can provide important information about the identity of the materials in each class. For example, the scatterplot of photoinfrared versus red bands shown here is useful for recognizing classes representing bare soil, vegetated areas, and water. You can view several spectral planes at the same time by opening more than one Ellipse ScatterPlot window.



Magnified view of the boxed area with pixel display turned on in addition to ellipse display. Black dots represent positions where points from different classes are superimposed.

A selected class ellipse is highlighted by solid color fill.

Keep the Ellipse ScatterPlot window open for the following exercises. You may want to minimize the window temporarily to conserve screen space.

Classification Dendrogram

The Classification Dendrogram is a tree-like plot that shows the degree of relatedness of the output classes. The dendrogram process performs a successive grouping of pairs of classes, beginning with the pair with the closest class centers in spectral space. As each pair of classes is merged, a new joint class center is computed and class-center distances are recalculated. This process repeats until all classes have been merged into a single class.

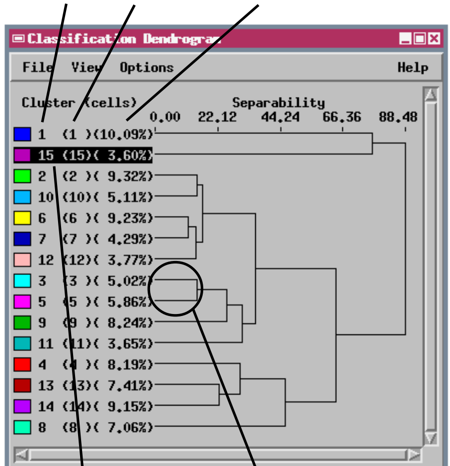
STEPS

☑ select Dendrogram from the Classification window's View menu

☑ click the selection button for Class 15 on the Operations on Classes window to deselect it

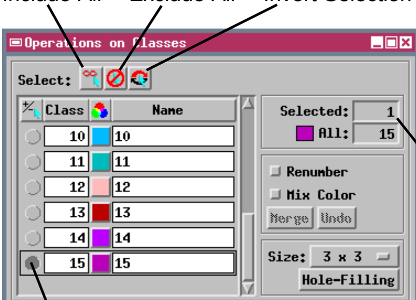
The dendrogram shows the classes merging from left to right, with the horizontal Separability axis representing distance in spectral space. The vertical lines joining two classes are plotted at the distance that separated the corresponding class centers before the classes were combined. Class pairs that join together near the left edge of the diagram are closely related in their spectral properties, and the degree of relatedness decreases to the right. In this example, classes 6 and 7 are most similar, having the lowest separability.

Class Class Name Class Size



Classes with similar spectral properties join near the left side of the Classification Dendrogram.

Include All Exclude All Invert Selection



A selected class (in this case the one selected in the previous exercise) is highlighted in black in the Dendrogram window.

The Selected field shows the number of classes currently selected. The All field shows the total number of classes.

A selected class is marked by a dark gray button in the Operations on Classes window. You can use these buttons to select or deselect individual classes, or select classes from the View window or any of the analysis tool windows.

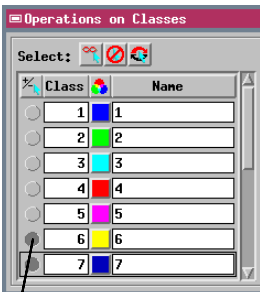
Keep the Classification Dendrogram window open for the following exercises.

Co-occurrence Analysis

STEPS

☑ select Co-occurrence from the Classification window's View menu

☑ select classes 6 and 7 in the Operations on Classes window

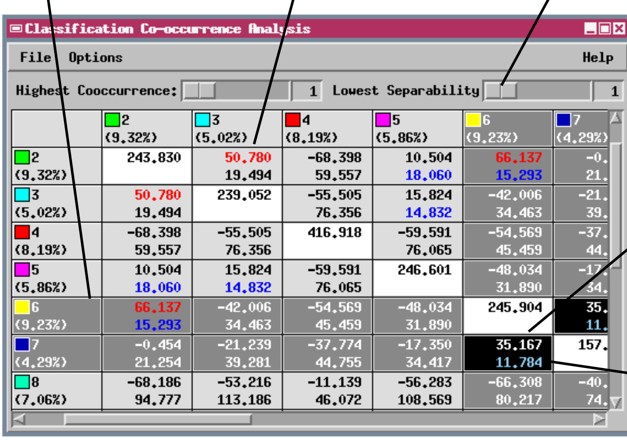


When you select classes in one of the other Classification windows, the Co-occurrence Analysis window automatically scrolls to show the values for the selected classes, and changes the background color for their rows and columns in the matrix.

The co-occurrence procedure analyzes the spatial associations of pairs of classes. It determines the frequency with which cells of each class pair occur adjacent to each other in the image. These values allow you to judge which classes are spatially associated.

The Co-occurrence Analysis window shows a matrix with both the co-occurrence value (upper number) and the separability value (lower number) for each pair of classes. Normalized co-occurrence values are shown by default. These values are produced by comparing the raw frequencies of adjacency with the values expected from a random distribution of class cells, a calculation that removes bias related to differing class sizes. A positive value indicates that two classes are adjacent to each other more often than random chance would predict. A negative value indicates that two classes tend not to occur together.

Matrix cell entries are shown in color for the 10 highest co-occurrence values and the 10 lowest separability values. You can use the sliders at the top of the window to select one of the 10 rank levels; the display automatically scrolls to the corresponding class pair and outlines the matrix cell in color. You can set text, cell, and highlight colors by choosing Colors... from the window's Options menu.



Co-occurrence value for classes 6 and 7

Separability value for classes 6 and 7

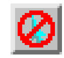
Keep the Co-occurrence Analysis window open for the following exercises.

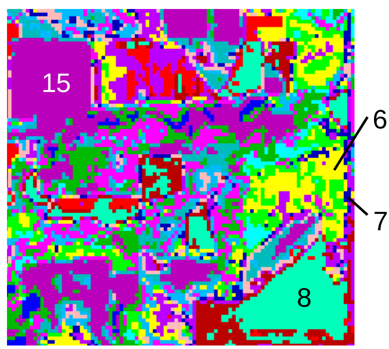
Analyzing Classes

You can use the set of analysis tools discussed in the preceding exercises, along with the class raster itself, to analyze the set of classes, determine class relationships, and identify pairs of classes for merging. Let's examine some examples from the ISODATA class raster you created for the Crow Butte TM scene.

As we noted on page 15, the dendrogram shows that classes 6 and 7 have very similar spectral properties. This similarity is also evident in the ellipse scatterplot, where their class ellipses overlap for all band combinations (choose some different band assignments for the plot axes to confirm this). The normalized co-occurrence value for classes 6 and 7 is high (35.167), showing that they also tend to be adjacent to each other in the scene. The class raster (see illustration at right) shows that class 7 commonly occurs as an "edge" class along the boundary between class 6 and other classes. Classes 6 and 7 seem to be good candidates for merging.

In contrast, classes 8 and 15 have dissimilar spectral properties, and their negative normalized co-occurrence value (-46.177) shows that they are not spatially intermixed. If we examine the input image (see illustration below right), we can see that Class 8 corresponds to bare agricultural fields with light-colored soil, while class 15 includes the "greenest" cultivated fields. Both classes occur in fairly uniform areas (fields) with little intermixing. It is evident that classes 8 and 15 have little in common spectrally or spatially, so they should remain as separate classes.

STEPS
- ☑ note the positions of the ellipses for classes 6 and 7 in the Ellipse Scatterplot
- ☑ click the Exclude All icon button on the Operations on Classes window to deselect all classes 
- ☑ select classes 8 and 15
- ☑ determine the Normalized Co-occurrence values for class pair 8 and 15, and check their positions in the Dendrogram and Ellipse Scatterplot



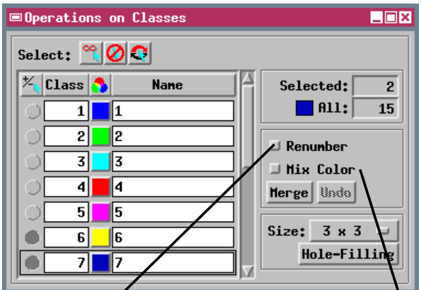
Portion of the ISODATA Class raster for Crow Butte illustrating spatial relationships of the selected classes.

"Natural" color image of the same area for comparison (TM Red, Green and Blue bands assigned to R, G, and B channels, respectively)

Merging Classes

STEPS
- ☑ click the Exclude All icon button on the Operations on Classes window
- ☑ select Classes 6 and 7
- ☑ click [Merge]

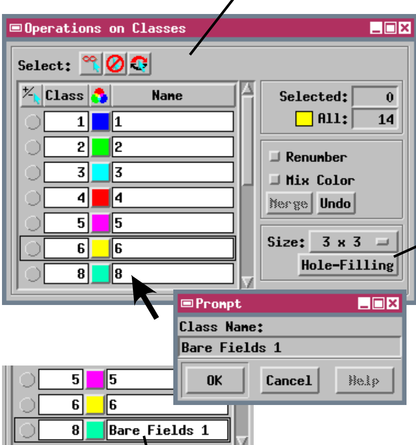
The Operations on Classes window provides controls that allow you to merge two or more classes. The merged class is assigned the number of the first selected component class, and the class raster view and all open class analysis windows are automatically updated. You can perform a number of merge operations, and if you are not satisfied with the result, you can press the Undo button repeatedly to return to any stage in the sequence. The final class raster is saved automatically when you exit the process. You can also use the Save As option on the Classification window's File menu to save any intermediate class results.



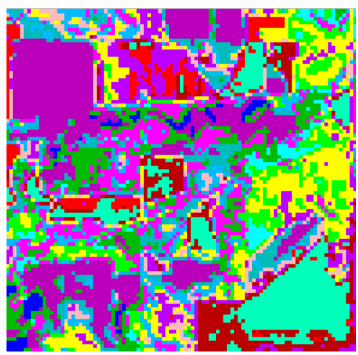
Turn on the Renumber option if you want to have the classes renumbered to remove gaps in the set of class numbers.

The Mix Color option creates a new custom class color for the merged class by mixing the colors of the input classes.

Operations on Classes window after merging class 7 with class 6.



Click in the Name field to open a prompt window and rename a class.



Part of updated class raster after merging class 7 with class 6. Compare with illustration on the previous page.

You can use the Hole-Filling option to simplify the class raster by applying a modal filter. The filter replaces the class of each cell with the class of the majority of cells in the surrounding neighborhood, removing isolated class cells. You can choose the size of the square filter window from the Size option menu.

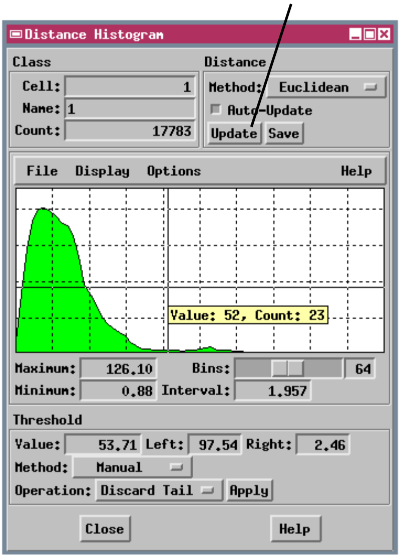
- ☑ Close the Co-occurrence Analysis, Classification Dendrogram, and Ellipse Scatterplot windows by choosing Close from each window's File menu.

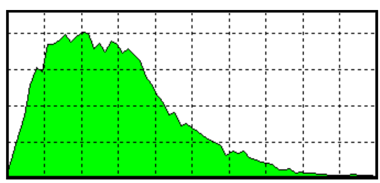
Distance Histogram

The Distance Histogram window can help you to assess the distribution of class points in spectral space for any class. It displays a distance histogram for the last selected class, derived from a distance raster that is automatically created for the current set of classes (see page 5 for a discussion of distance rasters). A diffuse class with many points far from the class center has a histogram with a "tail" extending to higher distance values. These outlier cells may represent distinctly different materials than cells near the class center. You can choose to remove the outlier cells from the class by using the histogram crosshairs to set a distance threshold and discard points at greater distances. The discarded cells are assigned a 0 value in the class raster.

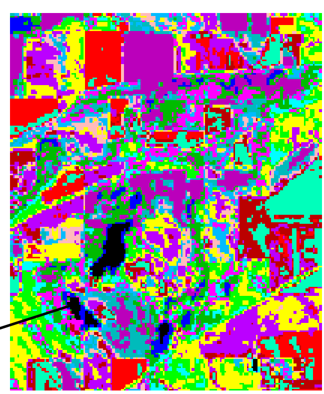
STEPS
- ☑ select Distance Histogram from the View menu on the Automatic Classification window
- ☑ in the Distance Histogram window, make sure that the Auto-Update toggle button in the Distance controls is turned on
- ☑ select class 1 in the Operations on Classes window
- ☑ move the Bins slider in the Distance Histogram window to set the number of histogram bins to 64
- ☑ click in the histogram pane, and drag the crosshair until the the Value ToolTip reads 52
- ☑ click the Apply button near the bottom of the window

If the Auto-Update option is turned off, you must press the Update button to recompute and display the distance histogram when you select a new class or change the number of bins for the current class.





New distance histogram for the smaller class 1. Note that there is no Undo for this operation!



Upper left portion of the ISODATA class raster for Crow Butte after discarding the tail of the class 1 histogram. Black cells were reassigned from class 1 to no class (cell value 0) by this operation.

Supervised Classification

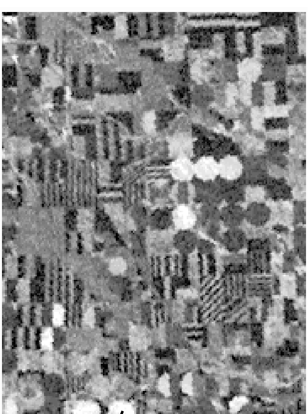
In **Supervised Classification**, you provide a **Training Set** raster which identifies representative sample areas for each of the desired output classes. The process determines the statistical properties of each of the training classes, then uses these properties to classify the entire image. Most of the methods assign all image cells to one of the training classes.

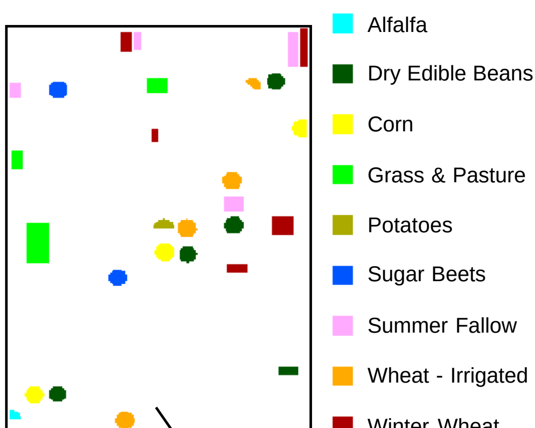
STEPS

- ☑ use the Display process to view each of the MNDVI raster objects in the BERMNDVI Project File (in the BEREA data collection)
- ☑ choose Display / Open in the Display Manager
- ☑ select the TS_LAYOUT object in the BEREATRN Project File to view the Training Set raster and legend

Now we move on to the Supervised Classification methods. In order to perform a supervised classification, you must have detailed knowledge of a portion of the study area, so that you can designate sample areas for each of the desired output classes that will be used to “train” the classification algorithm. The sample areas are incorporated in a special training set raster. The training set should incorporate as much of the spectral variability in the scene as possible.

The following classification exercises use a set of Modified Normalized Difference Vegetation Index (MNDVI) rasters for the Berea Creek West Quadrangle in western Nebraska — an agricultural area with both dryland and irrigated crops. Each MNDVI raster depicts variations in the density of vegetative cover for a different date during the 1981 growing season (May to October). The MNDVI rasters were derived from Landsat MSS data in the BEREAMSS Project File. Each value in an MNDVI raster is the scaled ratio of MSS band 7 (near infrared) to the sum of band 7 and band 5 (red), with corrections for path radiance (haze) in each band.





MNDVI raster for the Berea Creek West area for June 24, 1981. Brighter tones indicate greater cover by green vegetation.

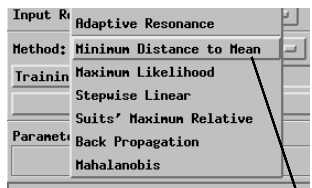
The Training Set raster for the Berea Creek West area identifies the major crop types for the 1981 crop cycle.

Minimum Distance Classification

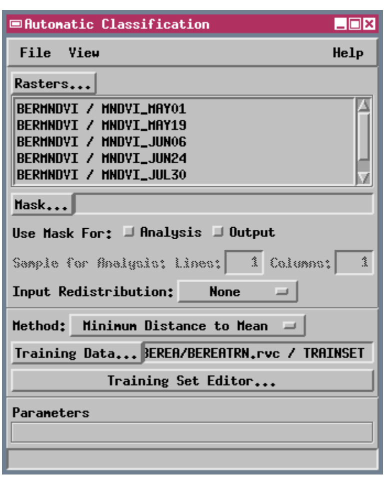
The training data you use in these exercises was derived from a 1981 vector crop map in the BERCROPS Project File. Many of the crop polygons in this object include more than one crop; the remaining single-crop polygons were used to create the training set raster used here. Later exercises show you how to use the Training Set Editor to create training set rasters for use in supervised classification.

As a first example of supervised classification, apply the Minimum Distance to Mean classifier to the Berea Creek MNDVI raster set. This method first analyzes the areas designated in the training set raster, then calculates a mean value in each band for each training class. These mean values define the location of the class center in spectral space. The process then assigns each cell in the input raster set to the class with the closest class center in spectral space.

STEPS
- ☑ from the BERMNDVI Project File, select the six MNDVI raster objects for classification
- ☑ select Minimum Distance to Mean from the Method option menu
- ☑ click [Training Data...] and select raster object TRAINSET from the BEREATRN Project File
- ☑ run the classification process, and place the output class raster in a new BEREASUP Project File



Classification methods below the horizontal divider in the Method option menu are supervised methods.



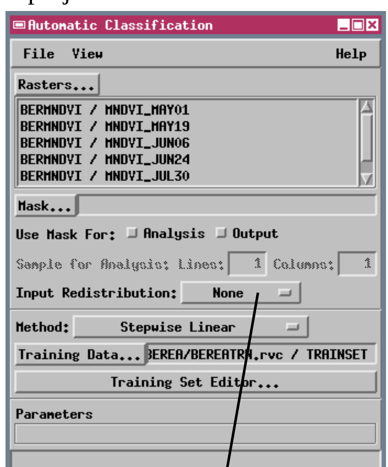
Class raster produced by the Minimum Distance method. Class colors are shown in the legend on the previous page.

Stepwise Linear Classification

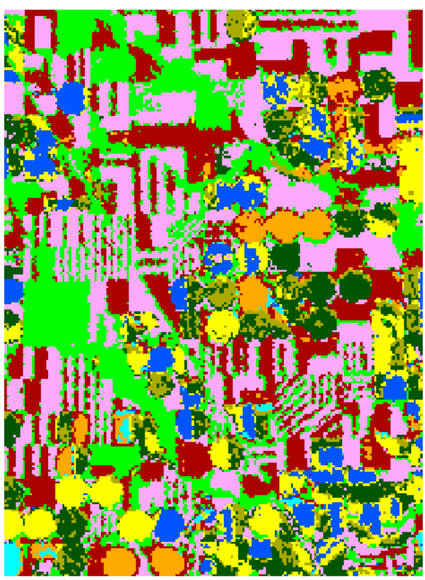
STEPS
☑ select Stepwise Linear from the Method option menu
☑ run the classification process and direct the output Class raster to the BEREASUP Project File

The Stepwise Linear method incorporates an automatic dimensional reduction procedure. Bands which do not add significantly to the discrimination of classes are eliminated from the classification process. This method is therefore particularly appropriate when your raster set includes a large number of input rasters.

The Stepwise Linear classification method applies techniques of linear discriminant analysis to classify image data. Discriminant analysis is a statistical technique that calculates a set of derived variables (discriminant functions) that are linear combinations of the bands in the input raster set. The discriminant functions define a new set of mutually-perpendicular coordinate axes in spectral space. The process analyzes the training set and chooses the set of discriminant functions that produces the best possible separation between the classes. Discriminant functions are chosen using a stepwise procedure that selectively adds and removes input bands to find the minimum number of bands necessary to produce the optimal separation of training classes. All input raster cells are then projected into the new coordinate system for classification.



Input redistribution is a pre-classification step that calculates new cell values for each input raster using procedures similar to the automatic contrast enhancement options in the Spatial Data Display process. You can choose from Linear, Normalize, Equalize, or Logarithmic methods. Input redistribution may provide better class discrimination for images that require contrast enhancement for effective viewing.



Class raster produced using the Stepwise Linear classification method.

Maximum Likelihood Classification

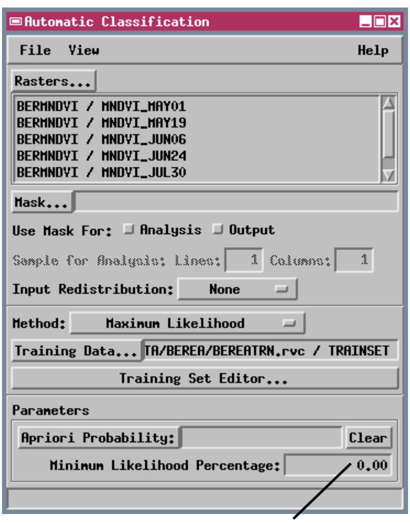
The Maximum Likelihood classification method applies probability theory to the classification task. From the training set classes the method determines the class centers and the variability in raster values in each input band for each class. This information allows the process to determine the probability that a given cell in the input raster set belongs to a particular training set class. The probability depends upon the distance from the cell to the class center, and the size and shape of the class in spectral space. The Maximum Likelihood method computes all of the class probabilities for each raster cell and assigns the cell to the class with the highest probability value. This method produces more accurate class assignments than the Minimum Distance to Means method when classes vary significantly in size and shape in spectral space.

STEPS

☑ select Maximum Likelihood from the Method option menu

☑ run the classification process, and direct the output Class raster to the BEREASUP Project File

Using imagery from multiple dates aids in discrimination of crop classes by taking advantage of the different planting, maturation, and harvest times of different crops. A time series of vegetation index rasters portrays variations in vegetation cover through time and enables the classification process to work with fewer input rasters compared to the raw MSS data for each date.



The Minimum Likelihood Percentage value is a threshold that you can set to exclude cells that don't fit any of the training classes particularly well. If the maximum class probability for a cell is less than this value, the cell is not classified, and is assigned a value of 0 in the class raster.

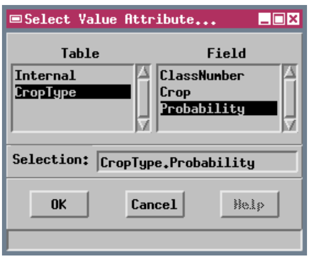
Class raster produced using the Maximum Likelihood classification method.

Using *A Priori* Probabilities

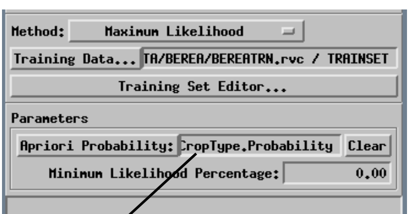
STEPS

☑ click [Apriori Probability] in the Automatic Classification window

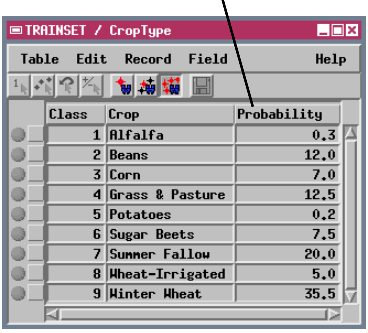
☑ in the Select Value Attribute window, select CropType from the Table list and Probability from the Field list, then click [OK]



☑ run the classification process, and direct the output Class raster to the BEREASUP Project File



The CropType table contains the *a priori* probability values assigned to each crop class for this exercise.



Class	Crop	Probability
1	Alfalfa	0.3
2	Beans	12.0
3	Corn	7.0
4	Grass & Pasture	12.5
5	Potatoes	0.2
6	Sugar Beets	7.5
7	Summer Fallow	20.0
8	Wheat-Irrigated	5.0
9	Winter Wheat	35.5

The probability values calculated by the Maximum Likelihood classifier in its default mode are based solely on spectral characteristics. But in some cases you may know independently that one class should be rare in the scene while another class should be very common. This prior knowledge could come from historical data (for example, records of the proportions of the area planted to different crops), or current information on similar areas. A probability value based on such information is termed an *a priori* probability.

You can incorporate *a priori* probabilities in the Maximum Likelihood classification process. The values must be in decimal form (between 0 and 1.0), and must be tabulated for each class in a single field in a database table attached to the training set raster. The *a priori* probability values are used as weighting coefficients in calculating class assignment probabilities.

The Maximum Likelihood class raster produced using *a priori* probabilities has fewer cells assigned to the rare class Potatoes.

View the Training Set Error Matrix

Selecting the Error Matrix option launches a classification error analysis, which uses a ground truth raster with sample areas of known class to evaluate the accuracy of the current class raster. The class of each sample area cell is compared to the class assigned to the corresponding cell in the class raster. The results are shown in the Error Matrix.

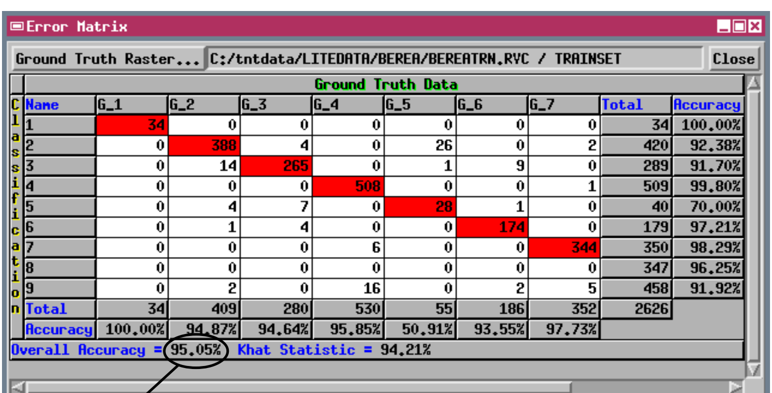
Each row in the Error Matrix represents an output class and each column a ground truth class. The value in each matrix cell is the number of pixels (raster cells) with the corresponding combination of output class and ground truth class. For each cell on the leading diagonal of the matrix (highlighted in color), the output class equals the input class, so the values in these cells give the number of correctly classified pixels for each class. The values in off-diagonal matrix cells represent incorrectly classified pixels. The Overall Accuracy value is calculated by dividing the total number of correctly classified raster cells (the sum of the leading diagonal values) by the total number of cells in the ground truth raster, and expressing the result as a percentage.

STEPS

☑ select Error Matrix from the Automatic Classification window's View menu

☑ click [Ground Truth Raster] in the Error Matrix window and select object TRAINSET from the BEREATRN Project File

Keep in mind that the Error Matrix shows classification accuracy only relative to the set of classes that you provide. Low accuracy values for particular classes may indicate that the sample areas you used are not completely representative of the class, that the class is not sufficiently different from other classes in its spectral properties, or that your set of classes does not include all of the significant materials in the scene.



Error Matrix

Ground Truth Raster... C:/tntdata/LITEDATA/BEREA/BEREATRN.RVC / TRAINSET Close

Ground Truth Data / Classification

Name	G_1	G_2	G_3	G_4	G_5	G_6	G_7	Total	Accuracy
1	34	0	0	0	0	0	0	34	100.00%
2	0	388	4	0	26	0	2	420	92.38%
3	0	14	265	0	1	9	0	289	91.70%
4	0	0	0	508	0	0	1	509	99.80%
5	0	4	7	0	28	1	0	40	70.00%
6	0	1	4	0	0	174	0	179	97.21%
7	0	0	0	6	0	0	344	350	98.29%
8	0	0	0	0	0	0	0	347	96.25%
9	0	2	0	16	0	2	5	458	91.92%
Total	34	409	280	530	55	186	352	2626	
Accuracy	100.00%	94.87%	94.64%	95.85%	50.91%	93.55%	97.73%		

Overall Accuracy = 95.05% Khat Statistic = 94.21%

The Overall Accuracy value for the training set raster in this example means that 95.05% of the sample area cells in the training set raster were correctly classified by the Maximum Likelihood classifier.

Keep the Error Matrix window open for the next exercise.

View a Ground Truth Error Matrix

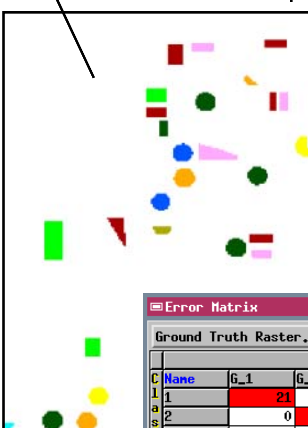
STEPS

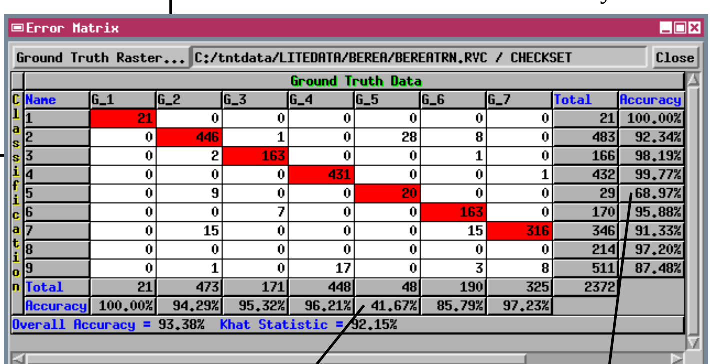
☑ click [Ground Truth Raster] in the Error Matrix window and select object CHECKSET from the BEREATRN Project File

Because the raster cells in training areas are used to "train" a supervised classifier, classification accuracy is usually higher for these sample cells than for other areas in the scene. To get a better idea of the broader classification accuracy, you can use a second set of ground truth areas that were not used in the training set. The CHECKSET raster object contains such a duplicate set of ground truth areas for the Berea Creek dataset. Using this ground truth raster, the Error Matrix shows an overall accuracy of 93.38%, somewhat less than that for the training set raster.

The Error Matrix shows two measures of accuracy for individual classes. The accuracy values for each column indicate the percentage of cells in that ground truth class that were correctly classified. Values less than 100% indicate errors of omission (ground truth cells omitted from the output class). This value is sometimes called the producer's accuracy. Conversely, the accuracy values for each row show the percentage of sample cells in each output class that were correctly classified. Values less than 100% indicate errors of commision (cells incorrectly included in the output class). This value is sometimes termed the user's accuracy.

Ground truth raster CHECKSET.





Error Matrix

Ground Truth Raster... C:/tntdata/LITEDATA/BEREA/BEREATRN.RVC / CHECKSET Close

Ground Truth Data

Classification Name	G_1	G_2	G_3	G_4	G_5	G_6	G_7	Total	Accuracy
1	21	0	0	0	0	0	0	21	100.00%
2	0	446	1	0	28	8	0	483	92.34%
3	0	2	163	0	0	1	0	166	98.19%
4	0	0	0	431	0	0	1	432	99.77%
5	0	9	0	0	20	0	0	29	68.97%
6	0	0	7	0	0	163	0	170	95.88%
7	0	15	0	0	0	15	316	346	91.33%
8	0	0	0	0	0	0	0	214	97.20%
9	0	1	0	17	0	3	8	511	87.48%
Total	21	473	171	448	48	190	325	2372	
Accuracy	100.00%	94.29%	95.32%	96.21%	41.67%	85.79%	97.23%		

Overall Accuracy = 93.38% Khat Statistic = 92.15%

Class 5 shows the lowest classification accuracies for the Berea Creek dataset. Its producer's accuracy is only 41.67% (only 20 of the 48 cells with this ground truth class correctly classified). Its user's accuracy is considerably higher, 68.97% (20 of the 29 cells in this output class correctly classified).

☑ click [Close] on the Error Matrix window

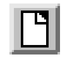
Make a Training Set from Vector Polygons

The Training Set Editor in the Automatic Classification process provides a flexible, streamlined interface for preparing or editing training data. When you click the New Training Set icon button, the editor creates a blank training set raster with the same raster dimensions and georeferencing as the rasters selected for classification. The editor allows you to create a training set raster in several ways, including using vector polygons as a source.

STEPS

☑ click [Training Set Editor...] on the Automatic Classification window

☑ click the New Training Set icon button on the Training Set Editor window



☑ click [Source] and select object TRAINVEC from the BEREATRN Project File

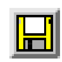
☑ select By Attribute from the Class option button menu

☑ click [Attribute:]

☑ in the Select Attribute Field window, select CropType from the Table list and Crop from the Field list, then click [OK]

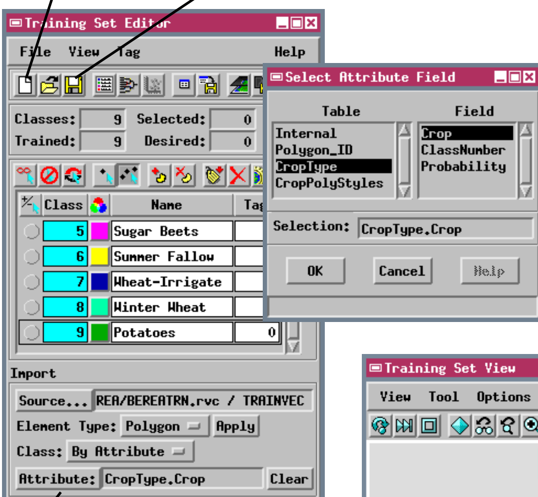
☑ click [Apply]

☑ click the Save Training Set icon button, and save the raster in a new Project File

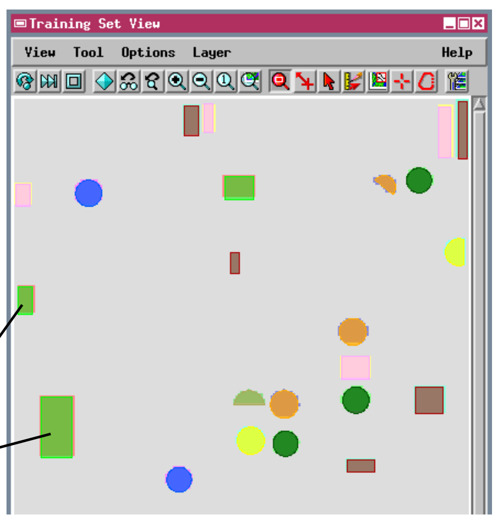


New Training Set

Save Training Set



The controls on the Import panel allow you to select a vector object source and to specify a database table and field containing the class names. You can use several source objects in sequence to build the training set raster.

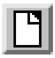


A training set raster is a categorical raster in which each training area is a collection of raster cells with a unique, nonzero value. A single training class can therefore include several noncontiguous subareas.

☑ choose Close from the File menu on the Training Set Editor window

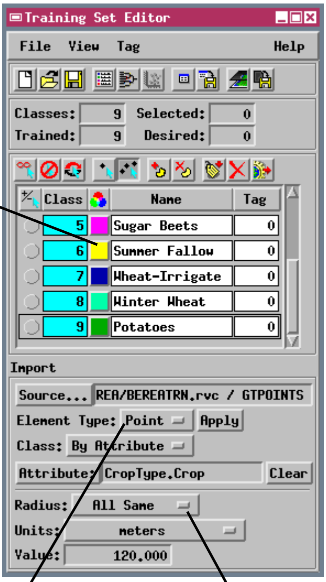
Create a Training Set from Vector Points

STEPS

- ☑ reopen the Training Set Editor
- ☑ click the New Training Set icon button on the Training Set Editor window 
- ☑ click [Source] and select object GTPOINTS from the BEREATRN Project File
- ☑ select By Attribute from the Class option button menu
- ☑ click [Attribute]
- ☑ in the Select Attribute Field window, select CropType from the Table list and Crop from the Field list, then click [OK]
- ☑ select All Same from the Radius option button menu
- ☑ enter 120 in the Value field
- ☑ click [Apply]

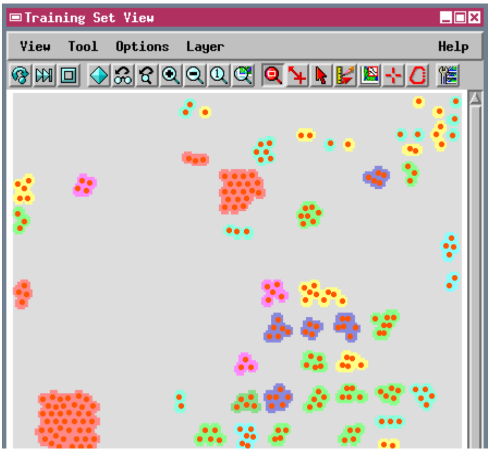
You can also use points in a vector object or database pinmap as a source of training data. These points might represent field observations collected at preselected random locations using a GPS unit to verify position coordinates. Each point generates a circular training area for which you must specify a radius. You can set a single radius for all points (as in this exercise), or apply a unique radius for each class from a field in a database table.

When you import classes, class colors are assigned internally and do not default to colors you may have assigned to vector point or polygon classes for display by attribute. You can click on the class color sample to open a standard Color Editor window and change any class color.



If the selected vector object includes both points and polygons, use the Element Type option button to select which type of element to use.

Use the All Same option and the Value field to set a single radius for all points. Use the By Attribute option to specify a database field containing class-specific radius values.

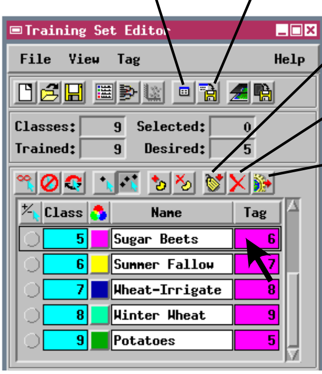


Keep the Training Set Editor open for the next exercise.

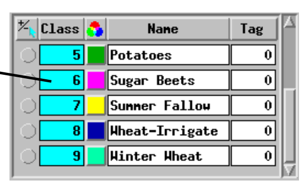
Use Tags to Renumber Training Classes

If the imported classes do not have the desired class numbers, you can use the Tag field in the class list to assign new raster values for the classes. Tags are temporary values that can be changed or cleared until you click the Apply Tag button, which permanently changes the class numbers. You can assign tag values for some or all of the classes before applying. If you assign the same tag value to more than one class, those classes will be merged. All classes to be combined must be tagged; tag values that duplicate existing, untagged class numbers are ignored when tags are applied.

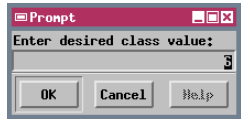
The Save Class Table option saves the current set of class numbers, names, and associated colors in a specialized database table which can be reopened later in the Training Set Editor. The table can be saved in a main-level database object or in a database attached to the current training set raster. These options allow you to make and use a consistent set of classes and colors for related datasets.

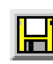


Class list after applying the tags shown at left



STEPS
- ☑ click in the Tag field for the Sugar Beets class
- ☑ enter the number 6 in the Prompt window and press [OK]



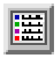
- ☑ repeat the previous steps to assign a tag value of 7 to Summer Fallow, 8 to Wheat-Irrigated, 9 to Winter Wheat, and 5 to Potatoes
- ☑ click [Apply Tag] 
- ☑ click [Save Class Table] and save the database object with the Class Information Table in the same Project File as your training set rasters 
- ☑ click the Save Training Set icon button, and save the raster in your training set Project File 

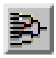
You can open a previously-saved Class Information Table prior to importing training areas from another object. You can still assign training areas to classes using the By Attribute option, but be sure that the class names in the designated database field match the names in the Class Information Table. The import procedure assigns training areas to your class list by matching these fields.

Keep the Training Set Editor open for the next exercise.

Training Statistics and Dendrogram

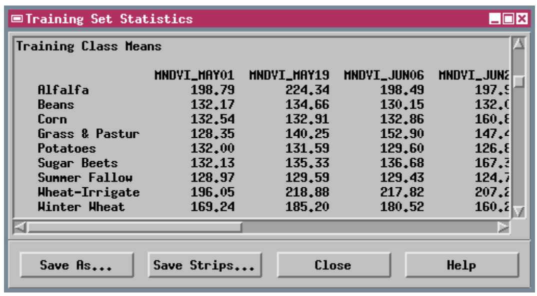
STEPS



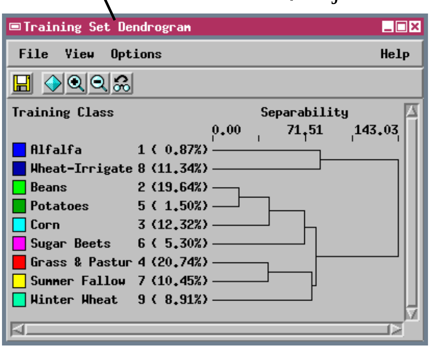


- ☑ click on the Statistics icon button
- ☑ click on the Dendrogram icon button
- ☑ close the Training Set Statistics window and the Training Set Dendrogram window after you have examined them

The Training Set Editor provides several tools to help you evaluate your training set raster before using it in the classification procedure. The Training Set Statistics and Training Set Dendrogram provide the same information for the training set as the Output Statistics and Classification Dendrogram provide for a class raster (see pages 13 and 15). You can use this information to judge the spectral characteristics and separability of your training classes.



Use the Training Set Dendrogram to help evaluate the suitability of your training classes.



Objects that you are using as a source for importing training areas must be georeferenced, but need not match the extents of the training set raster. The Training Set Editor automatically registers the source object to the training set raster, and transfers training areas only in the overlap area. The source object can even be another raster object with different extents and cell size (as long as it is an unsigned 4-, 8-, or 16-bit raster). Classes are assigned by raster value, and you can use the By Attribute option to specify a raster database table containing class names. For example, you might have an image set and training set raster for a large area, but want to run a supervised classification on a higher-resolution image covering only a portion of the area. You could import training areas and the class list from the existing training set raster to ensure consistency between the class rasters.

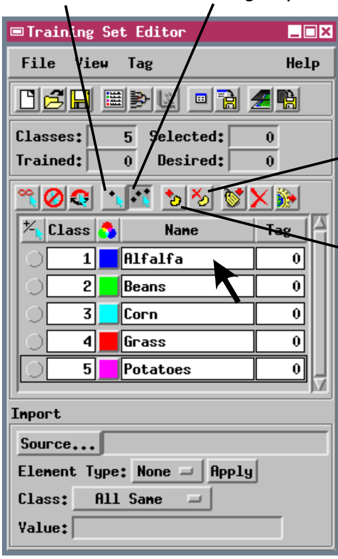
Create Training Classes Manually

The Training Set Editor also allows you to create a new training set raster manually by adding and naming classes and drawing training areas over a reference image, such as a scanned and georeferenced ground truth map. The first step is to set up a list of classes. After you add classes to the list, you can name and select classes in the Training Set Editor window in the same way as in the Operations on Classes window (see pages 15 and 18). The Training Set Editor window allows you to set either single selection or multiple selection modes for selecting classes from the class list.

STEPS
- ☑ reopen the Training Set Editor
- ☑ select Controls from the Layer menu in the Training Set View window
- ☑ in the Layer Manager window, click the Add Raster icon button and choose Single from the dropdown menu 
- ☑ select object REFERENCE from the BEREATRN Project File
- ☑ click the New Training Set icon button on the Training Set Editor window 
- ☑ click the Add Class icon button on the Training Set Editor window five times to add five classes 
- ☑ click in the Name field for the first class, enter the name Alfalfa in the popup Prompt window, and click [OK]
- ☑ repeat for the next four classes, entering the names shown in the illustration

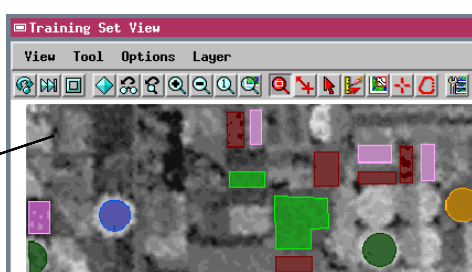
Single Selection

Multiple Selection mode is useful for group class operations such as merging, tagging, deleting, and releasing.



Delete Selected Class

Add Class



The reference layer for this exercise is a grayscale satellite image of the Berea area with ground truth areas indicated in color. You would most likely use a scanned topographic or planimetric map or aerial photograph with hand-drawn training areas.

This exercise continues on the next page.

Draw a Training Area

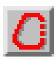
STEPS

☑ use the View window controls to zoom in on the green fields near the top of the reference image

☑ click the Single Selection icon button on the Training Set Editor window

☑ click the selection button for Class 2 (Beans)

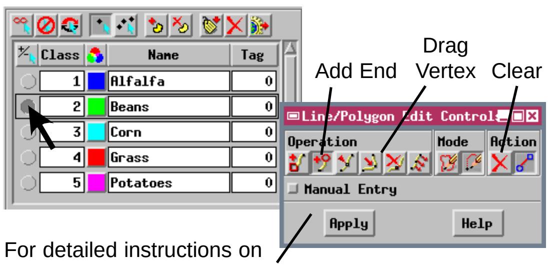
☑ click the Select Area icon button on the Training Set View window



☑ use the Line / Polygon Edit Controls to outline the circular dark green field (1) as shown (keeping well inside the edge)

☑ right-click and select Assign All Cells from the popup menu

In order to draw a training area, you must first select a single class from the list in the Training Set Editor window. Any areas you draw are associated with the currently selected class. The Select Area icon button on the Training Set View window opens a standard Line / Polygon Edit Controls window. These controls allow you to draw simple or complex polygonal shapes outlining the training areas. After you draw a polygon, you must choose an assignment option from the right mouse button menu before proceeding to draw additional polygons.



For detailed instructions on using the Line / Polygon Edit Controls, consult the Getting Started booklet *Editing Vector Geodata*.

This exercise continues on the next page.

After drawing a polygon, click the right mouse button to show the popup menu. (Don’t use the Apply button). The assignment options make it easy to work with training areas of different classes that are close together.

Assign Free Cells: all cells in the polygon that do not yet have a class assignment are added to the selected training class.

Assign All Cells: all cells in the polygon are added to the selected training class regardless of previous class assignment.

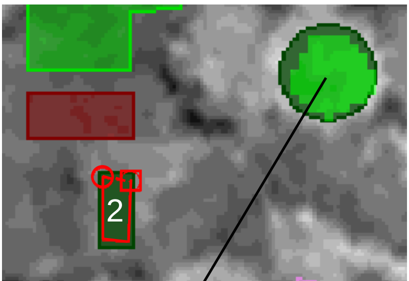
Release All Cells: all cells in the polygon are marked as unclassified.

Release Selected: all cells of the selected class (or classes) are marked as unclassified.

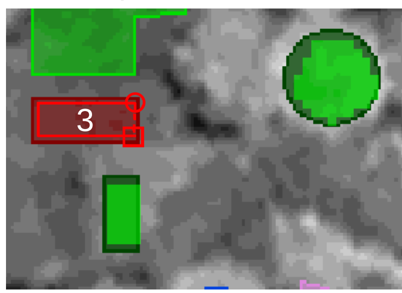
Release Class Assignments

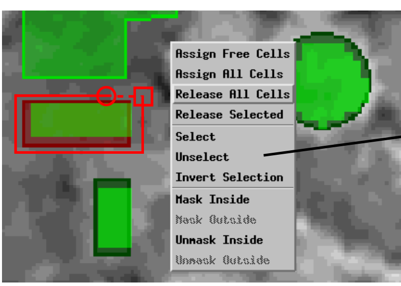
If you assign a training area to the wrong class, or make the area too large, you can also use the Select Area tool to correct these problems. To “undo” a previous class assignment and return the area to an unclassified status, trace around the area and use one of the Release options on the right mouse button menu.

STEPS
- ☑ repeat the last two steps for the rectangular dark green field (2)
- ☑ repeat again for the dark brown rectangle (3) to intentionally misassign this area to class 2
- ☑ draw another polygon completely enclosing the incorrect field 3 training area
- ☑ right-click and select Release All Cells from the popup menu
- ☑ select Class 4 (Grass)
- ☑ draw a polygon outlining the bright green field (4)
- ☑ right-click and select Assign All Cells from the popup menu

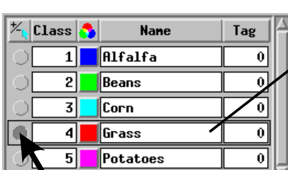


Field 1 is drawn in bright green after assignment to class 2. The training set raster is displayed with partial transparency over any background reference objects.

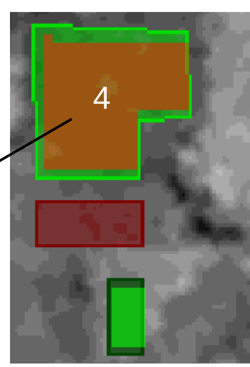




You can also use the Select Area tool to perform class selection operations for any class with a training area.



You can add training areas in any order. Just be sure that you select the appropriate class before drawing the desired area.

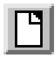


Keep the Training Set Editor open for the next exercise.

Make a Training Set Mask

STEPS

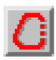
☑ click the New Training Set icon button on the Training Set Editor window



☑ click [No] in the Verify window that asks whether you want to save the training set raster

☑ use the View window controls to pan to the white polygon in the northwest quadrant of the reference image

☑ click the Select Area icon button on the Training Set View window



☑ use the Line / Polygon Edit Controls to outline the white polygon as shown

☑ right-click and select Mask Inside from the popup menu

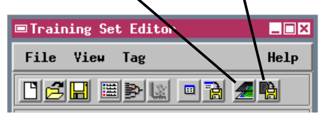
☑ click the Save Mask icon button on the Training Set Editor window and save the raster in your training set Project File

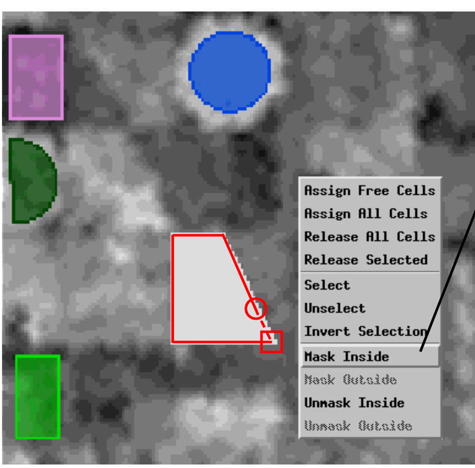
The manual drawing functions in the Training Set Editor also allow you to create a mask raster for use in classification (see page 9). You can draw over a reference image and define areas to include or exclude from the mask. The raster you make in this exercise masks the area of a small town, which you would probably want to exclude from a classification focusing on crops.

After selecting the desired reference image for display in the Training Set View window, open a new training set raster, which will become the mask. Choose the Select Area tool to open the Line / Polygon Edit Controls window, and draw polygonal areas as in the previous exercises. Use the mask options on the right mouse button menu to either mask or unmask the area inside the polygon. (Additional mask options will be implemented in future versions.) Masked areas are displayed in a darker transparent gray in the View window. Use the Save Mask function to save the mask as a binary raster.

You can use the Open Mask function to select an existing binary raster for editing in the Training Set Editor.

Save Mask





Mask Inside: the area inside the polygon becomes part of the mask (assigned a value of 0 in the binary raster).

Unmask Inside: any previously masked area inside the polygon is removed from the mask (assigned a value of 1 in the binary raster).

These options can also be used to create and edit mask areas within a regular training set raster.

What Next?

Now that you have completed these exercises in automatic image classification, you are ready to begin experimenting with your own imagery. You need not confine the classification process to the types of multispectral optical imagery mentioned here. The thermal infrared band of Landsat Thematic Mapper imagery (Band 6) can add a useful dimension to the classification process in some applications. The recent advent of commercial satellite radar imagery also makes it possible to combine optical and radar imagery covering the same area, providing higher levels of discrimination of surface feature types. Keep in mind, however, that all rasters in an input raster set must have the same geographic extents and the same cell size. You can use the Automatic Raster Resampling process (Raster / Resample and Reproject / Automatic) to produce an appropriate coextensive raster set from overlapping georeferenced raster sets from different sources (see the tutorial booklet entitled *Rectifying Images* for more information).

Finding Additional Information

This booklet has provided only a brief introduction to image classification methods and concepts. The references listed below provide good places to start if you are interested in finding additional information about automatic image classification and individual classification methods.

Drury, Stephen A. (2001), Image Interpretation in Geology (3rd ed.). Chapter 5, Digital Image Processing. New York: Routledge. pp. 145-152.

Jensen, John R. (1996). *Introductory Digital Image Processing: a Remote Sensing Perspective* (2nd ed.). Chapter 8, Thematic Information Extraction: Image Classification. Upper Saddle River, NJ: Prentice-Hall. pp. 197-256.

Johnston, R. L. (1978). *Multivariate Statistical Analysis in Geography: A Primer on the General Linear Model*. Chapter 8, Discriminant Analysis. New York: Longman, Inc. pp. 234-252.

Lillesand, Thomas M. (2004). *Remote Sensing and Image Interpretation* (5th ed.). Chapter 7, Digital Image Processing. New York: John Wiley and Sons. pp. 550-610.

Schowengerdt, Robert A. (1997). *Remote Sensing: Models and Methods for Image Processing*. Chapter 9, Thematic Classification. New York: Academic Press. pp. 389-438.

Tou, Julius T. and Gonzales, Raphael C. (1974). *Pattern Recognition Principles*. Reading, MA: Addison-Wessley. 377 pp.

Index

MicroImages, Inc.

www.microimages.com